Predicting Pareto and Exponential Observables

by

Seymour Geisser

University of Minnesota

Technical Report No. 408

October 1982

PREDICTING PARETO AND EXPONENTIAL OBSERVABLES

by

Seymour Geisser[1]

University of Minnesota

1. Introduction

In a previous paper, Geisser (1982), we discussed the importance, in certain statistical analyses, of calculating a probability function for the fraction of a finite number of future observations that lie in some interval. Exact, approximate and asymptotic results for the probability, when the underlying distribution was a simple exponential, were also obtained.

In this paper we extend the exact prediction results to the translated exponential case and show that the results are appropriate for the Pareto case as well. In the next two sections, exact and approximate results are derived for the translated exponential under the typical "non-informative" prior distribution for the two parameters involved when censoring is of the type most often encountered in practice. The fourth section establishes the connection between the Pareto distribution and the translated exponential in terms of prediction. An example follows in the next section.

The Pareto distribution has been the focus of attention of a number of workers recently, e.g. Lwin (1974), Dyer (1981), Arnold and Press (1982). Lwin considered the estimation of the two Pareto parameters using a joint conjugate prior which Arnold and Press criticized. They in turn suggested another joint prior which they claimed was not susceptible to previous criticism. Dyer considered the problem of estimating the survival function in the uncensored case using the structural approach of Fraser (1979) which, in effect, yields a "non-informative" prior for the Pareto parameters. If perceived from the structural viewpoint, the results

1. Research supported by NIH grant GM25271.

derived here can be considered an extension of Dyer's results in two respects.

First the work is extended to estimating the survival function in the censored case and secondly, and more importantly, to the problem of predicting the fraction of a finite number of future values that lie in some set. The survival function is basically the limiting case. However, we prefer to regard this as basically a predictive application of a Bayesian approach.

2. Prediction for Translated Exponential Observables.

Let $X_1, \ldots, X_N$ be realizations from the translated exponential density

$$f(x|\alpha,\gamma) = \alpha e^{-\alpha(x-\gamma)} \quad \text{for } \alpha > 0,\ x > \gamma > -\infty \tag{2.1}$$

The distribution and survival functions are respectively

$$\begin{aligned} F(x|\alpha,\gamma) &= 1 - e^{-\alpha(x-\gamma)} && \text{for } x > \gamma \\ &= 0 && \text{otherwise} \end{aligned} \tag{2.2}$$

$$\begin{aligned} \Pr(X > x|\alpha,\gamma) &= e^{-\alpha(x-\gamma)} && \text{for } x > \gamma \\ &= 1 && \text{otherwise} \end{aligned} \tag{2.3}$$

Suppose that $x_1,\ldots,x_d$ are the fully observed values while $X_{d+1},\ldots,X_N$ are censored at $x_{d+1},\ldots,x_N$ respectively. Let $\bar{x}_d = d^{-1}(x_1+\ldots+x_d)$ and $m_d = \min(x_1,\ldots,x_d)$ and further order the censored values $x_{d+1},\ldots,x_{N+d}$ as follows: $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(N-d)}$.

Suppose that for some $k \in [0,N-d]$ the sample is such that

$$x_{(k+1)} < m_d < x_{(k)} \tag{2.4}$$

where $x_{(0)} = \infty$ and $x_{(N-d+1)} = -\infty$ then the likelihood is, for $\alpha > 0$

$$L(\alpha,\gamma) = \begin{cases} \alpha^d e^{-\alpha d(\bar{x}_d-\gamma)-\alpha k(\bar{x}_{(k)}-\gamma)} & \text{for } x_{(k+1)} < \gamma \leq m_d \\ \\ \alpha^d e^{-\alpha d(\bar{x}_d-\gamma)-\alpha j(\bar{x}_{(j)}-\gamma)} & \text{for } x_{(j+1)} < \gamma \leq x_{(j)} \end{cases} \tag{2.5}$$

where $j = k+1,\ldots,N-d$, and $\bar{x}_{(h)} = h^{-1}(x_{(1)}+\ldots+x_{(h)})$.

If we employ the usual vague joint prior density where $\log\alpha$ and γ are uniform,

$$p(\alpha,\gamma) \propto \alpha^{-1} \tag{2.6}$$

we obtain for $d \geq 2$ and $\underset{\sim}{x} = (x_1,\ldots,x_N)$

$$p(\alpha,\gamma \mid \underset{\sim}{x}) \propto \begin{cases} \alpha^{d-1} e^{-\alpha[d(\bar{x}_d-\gamma)+k(\bar{x}_{(k)}-\gamma)]} & x_{(k+1)} < \gamma \leq m_d \\ \alpha^{d-1} e^{-\alpha[d(\bar{x}_d-\gamma)+j(\bar{x}_{(j)}-\gamma)]} & x_{(j+1)} < \gamma \leq x_{(j)} \\ 0 \quad \text{elsewhere} & \end{cases} \tag{2.7}$$

Although there is no inherent difficulty in carrying through the computations that we require for the complete though messy solution using (2.7), we shall, at this stage, avoid this by making a simplifying but sensible assumption.[2] Assume that the sample is such that $x_{(N-d)} \geq m_d = m$ i.e. the minimum value, m_d, of the fully observed values is the minimum, m, of all values in the sample, fully observed and censored.

When this is the case we obtain

$$p(\gamma|\alpha) = N\alpha e^{\alpha N(\gamma-m)} \qquad \text{for } \gamma < m \tag{2.8}$$

$$p(\alpha) = [N(\bar{x}-m)]^{d-1} \frac{\alpha^{d-2} e^{-\alpha N(\bar{x}-m)}}{\Gamma(d-1)} \qquad \alpha > 0 \tag{2.9}$$

for $\bar{x} = N^{-1}[x_1+\ldots+x_N]$.

For a single future observable Z we can calculate, respectively, the predictive distribution function and survival function

$$F(z) = \Pr(Z \leq z) = E_{\gamma,\alpha}\,(1-e^{-\alpha(z-\gamma)}) \tag{2.10}$$

$$\Pr(Z>z) = 1 - F(z). \tag{2.11}$$

The latter result is

$$\Pr(Z > z) = \begin{cases} \dfrac{N^d(\bar{x}-m)^{d-1}}{(N+1)[z-m+N(\bar{x}-m)]^{d-1}} & z > m \\[2ex] 1 - (N+1)^{-1}\left(\dfrac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} & z \leq m \end{cases} \tag{2.12}$$

2. Arguments supporting this assumption appear in section 6.

Note that for $z = \bar{x}$ and $d = N$, the fully observed case

$$\Pr(Z > \bar{x}) = \left(\frac{N}{N+1}\right)^N \tag{2.13}$$

which yields the same as the (unfamiliar) sampling result i.e.

$$\Pr(Z > \bar{X}) = \left(\frac{N}{N+1}\right)^N \tag{2.14}$$

where Z and X_i are i.i.d. translated exponential random variables.[3] A familiar result (in sampling theory)

$$\Pr[Z > m] = \frac{N}{N+1} \tag{2.15}$$

also drops out as a special case of (2.12).

Consider now a set of future values $Z_1,\ldots,Z_M$ from (2.1) and the fraction of them that exceed a given value, z, say.

Let

$$Y_i = \begin{cases} 1 & \text{if } Z_i > z \\ 0 & \text{if } Z_i \leq z \end{cases} \tag{2.16}$$

for $i = 1,\ldots,M$, so that $\bar{Y} = M^{-1}[Y_1 + \ldots + Y_M]$ is the required fraction. Now the conditional probability

$$P[Z_i > z \mid \alpha,\gamma] = \theta = \min(e^{-\alpha(z-\gamma)}, 1), \tag{2.17}$$

so that the probability function of $\bar{Y}$ can be obtained from

$$\Pr\left[\bar{Y} = \frac{r}{M} \mid z\right] = \int \binom{M}{r} \theta^r (1-\theta)^{M-r} \, dP(\theta \mid \underset{\sim}{x}) \tag{2.18}$$

where $P(\theta \mid \underset{\sim}{x})$ is the posterior distribution of the random variable Θ, or directly from (2.8) and (2.9). The distribution of Θ conditional on α is easily calculated from (2.8) to be

$$\Pr[\Theta \leq \theta \mid \alpha] = \begin{cases} 0 & \theta \leq 0 \\ \theta^N e^{\alpha N(z-m)} & 0 < \theta < \min(e^{\alpha(m-z)}, 1) \\ 1 & \theta \geq 1 \end{cases} \tag{2.19}$$

3. An interesting conjecture is that for i.i.d. absolutely continuous random variables (2.14) holds for all N if and only if the variables are translated exponentials.

with

$$\Pr[\Theta = 1|\alpha] = \begin{cases} 1 - e^{\alpha N(z-m)} & \text{if } z < m \\ 0 & \text{if } z \geq m \end{cases} \tag{2.20}$$

By taking the expectation of the conditional distribution (density) of Θ with respect to the posterior distribution of α, given by (2.9), the unconditional posterior distribution (density) of Θ is obtained. For $0 < \theta < 1$ and

$$I_a(p) = \int_0^p \frac{e^{-u} u^{a-1}}{\Gamma(a)} \, du \tag{2.21}$$

we obtain

$$\Pr[\Theta \leq \theta|\underset{\sim}{x}] = \begin{cases} \theta^N \left(\dfrac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} & z \leq m \\ \theta^N \left(\dfrac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} I_{d-1}\left[N\left(\dfrac{\bar{x}-z}{m-z}\right)\log\theta\right] + 1 - I_{d-1}\left[N\left(\dfrac{\bar{x}-m}{m-z}\right)\log\theta\right] & m < z \neq \bar{x} \\ N^{d-1}\theta^N[-\log\theta]^{d-1}/\Gamma(d) + 1 - I_{d-1}(-N\log\theta) & z = \bar{x} \end{cases} \tag{2.22}$$

and

$$\Pr(\Theta = 1|\underset{\sim}{x}) = \begin{cases} 1 - \left(\dfrac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} & z < m \\ 0 & z \geq m \end{cases} \tag{2.23}$$

Dyer (1981), using the structural approach of Fraser (1979), which is equivalent to using the "vague" prior density employed here, derived the distribution of Θ for the uncensored case. His results are then obtainable by setting $d = N$. Note also that for the case $z = \bar{x}$, the distribution of θ is no different then when γ is known. The latter case is given by Geisser (1982).

The distribution of $\bar{Y}$ is then derivable either from (2.22) or from (2.8) and (2.9) and is

$$
\Pr(\bar{Y} = \tfrac{r}{M} | z) = \begin{cases} \left(\dfrac{\bar{x} - m}{\bar{x} - z}\right)^{d-1} \binom{N+r-1}{r} \Big/ \binom{N+M}{M} & r < M,\ z < m \\ 1 - \dfrac{M}{N+M} \left(\dfrac{\bar{x} - m}{\bar{x} - z}\right)^{d-1} & r = M,\ z < m \\ N \binom{M}{r} \sum\limits_{j=0}^{M-r} \binom{M-r}{j} \dfrac{(-1)^j}{(N+r+j)} \left(1 + \dfrac{(r+j)(z-m)}{N(\bar{x} - m)}\right)^{-(d-1)} & m < z \\ \binom{M}{r} \sum\limits_{j=0}^{M-r} \binom{M-r}{j} (-1)^j \left(1 + \dfrac{r+j}{N}\right)^{-d} & z = \bar{x} \end{cases} \tag{2.24}
$$

Now either directly from (2.24) or invoking de Finetti's theorem we find that $\bar{Y} \to \Theta$ with limiting distribution function obtained from the probability function given by (2.22) and (2.23).

Suppose we are actually interested in the fraction of $Z_1, \ldots, Z_M$ that lies in (z_1, z_2). The previous results are easily adapted to include such a situation since

$$
\Pr\left[\bar{Y} = \tfrac{r}{M} | z_1, z_2\right] = \Pr\left[\bar{Y} = \tfrac{r}{m} | z_1\right] - \Pr\left[\bar{Y} = \tfrac{r}{m} | z_2\right] \qquad \text{for } z_2 > z_1. \tag{2.25}
$$

Clearly any measurable set can, of course, be handled similarly. It is also clear that

$$
E(\bar{Y} | z) = E(\Theta) = \Pr[Z > z] \tag{2.26}
$$

the latter is evaluated in (2.12). Further setting $\Pr[Z > z] = q$

$$
\operatorname{var}(\bar{Y}) = q(1-q)M^{-1}[1 + \rho(M-1)] \tag{2.27}
$$

where

$$
\rho = [\Pr[Y_i = 1, Y_j = 1] - q^2] \big/ q(1-q) \tag{2.28}
$$

and independent of i and j for $i \neq j$

$$\Pr[Y_i = 1,\ Y_j = 1] = \begin{cases} \dfrac{N^d(\bar{x} - m)^{d-1}}{(N+2)[2(z-m) + N(\bar{x} - m)]^{d-1}} & z > m \\[2ex] 1 - \dfrac{2}{N+2}\left(\dfrac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} & z \leq m, \end{cases} \tag{2.29}$$

From the above, we also obtain by means of a limiting argument i.e. $(\bar{Y} \to \Theta)$ that

$$\mathrm{Var}(\Theta) = \rho q(1-q) \tag{2.30}$$

so that

$$\mathrm{Var}(\bar{Y}) = \mathrm{Var}(\Theta) + q(1-q)(1-\rho)M^{-1}. \tag{2.31}$$

The value q can also be considered as an estimator of θ (it has some optimal frequentist properties in that it minimizes a Kulback-Leibler divergence measure) but if it is used for more than one future observation to "estimate" the probability of a future fraction $\bar{Y}$ by assigning $\bar{Y}$ the binomial probability function

$$\binom{M}{r} q^r(1-q)^r , \tag{2.32}$$

some difficulties ensue.

In this case

$$E(\bar{Y}) = q$$
$$\mathrm{Var}\ (\bar{Y}) = M^{-1}q(1-q) , \tag{2.33}$$

so that $\bar{Y} \to q$ as M grows. This obviously is unacceptable. Although q is perfectly sensible for a single future Z, it is inadequate to assume that a set of future Z's, although conditionally i.i.d., can be analyzed by this "estimative" approach. In other words the transformed Y's, also conditionally i.i.d., cannot be used with the "estimate" as just i.i.d. Bernoulli variates--but must be analyzed as unconditionally exchangeable Bernoulli variates.

3. An Approximation

Letting G(u) represent the distribution function of a χ^2 variate with 2d-2 degrees of freedom, we have from (2.22) for $m < z < \bar{x}$

$$\Pr[\Theta \leq \theta | \underset{\sim}{x}] = \theta^N \left(\frac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} G\left(2N\left(\frac{\bar{x}-z}{m-z}\right)\log\theta\right) + 1 - G\left(2N\left(\frac{\bar{x}-m}{m-z}\right)\log\theta\right) \tag{3.1}$$

and for $z > \bar{x}$

$$\Pr[\Theta \leq \theta | \underset{\sim}{x}] = \theta^N \left(\frac{\bar{x}-m}{z-\bar{x}}\right)^{d-1} G\left(2N\left(\frac{z-\bar{x}}{m-z}\right)\log\theta\right) + 1 - G\left(2N\left(\frac{\bar{x}-m}{m-z}\right)\log\theta\right) \tag{3.2}$$

Setting $\theta = \frac{r}{M}$, this is the limiting value, as M grows, for $\Pr[\bar{Y} \leq \frac{r}{M} | z]$. An accurate approximation for finite values of M for this probability is desireable because for moderately large M the exact value becomes more and more burdensome to compute. Previous work by Geisser (1982) has shown that for the γ known case the asymptotic result analogous to (3.2) even when θ is set equal to $(r+\frac{1}{2})/M$ to correct for continuity left much to be desired when used as an approximation for $\Pr[\bar{Y} \leq \frac{r}{M} | z]$ even for values of M as large as 100. In that case an approximation based on an F-variate was introduced and was a considerable improvement over the asymptotic result. With this in mind we offer as an analogous approximation; for $m < z < \bar{x}$

$$\Pr(\bar{Y} \leq \frac{r}{M} | z) \doteq \theta^N \left(\frac{\bar{x}-m}{\bar{x}-z}\right)^{d-1} H\left(\frac{N(\bar{x}-z)}{(d-1)(m-z)}\log\theta\right) + 1 - H\left(\frac{N(\bar{x}-m)}{(d-1)(m-z)}\log\theta\right), \tag{3.3}$$

and for $z > \bar{x}$

$$\Pr[\bar{Y} \leq \frac{r}{M} | z] \doteq \theta^N \left(\frac{\bar{x}-m}{z-\bar{x}}\right)^{d-1} H\left(\frac{N(z-\bar{x})}{(d-1)(m-z)}\log\theta\right) + 1 - H\left(\frac{N(\bar{x}-m)}{(d-1)(m-z)}\log\theta\right) \tag{3.4}$$

where H is the distribution function of an F-variate with 2d-2 and 2M degrees of freedom and θ is set equal to $(r+\frac{1}{2})/M$ to correct for continuity. For $z = \bar{x}$

$$\Pr[\bar{Y} \leq \frac{r}{M} | \bar{x}] \doteq 1 - H^*\left(-\frac{N}{d}\log\frac{r+\frac{1}{2}}{M}\right) \tag{3.5}$$

where H^* is the distribution function of an F-variate with 2d and 2M degrees of freedom.

For $z \leq m$, the exact calculation can be made for fairly large M and beyond that the asymptotic approximation should be adequate.

A rationale for the F-approximation is given in Geisser (1982) for the $\gamma \equiv 0$ case. The one here follows along similar lines though slightly more involved. It is based on the fact that if U is the average of M future Z's then the random variable

$$\frac{N(\bar{x}-m)}{(d-1)(U-m)}, \tag{3.6}$$

when calculated from the predictive distribution of U which is conditioned on $\bar{x}$ and m, has a density proportional to that of an $F_{2d-2,2M}$ variate when $U > m$.

4. Pareto Prediction

Suppose $W_1,\ldots,W_N$ is a random sample from the pareto distribution

$$F(w|\alpha,\beta) = \begin{cases} 0 & w<\beta \\ 1-\left(\frac{\beta}{w}\right)^{\alpha} & w>\beta>0 \end{cases} \tag{4.1}$$

and density

$$f(w|\alpha,\beta) = \frac{\alpha\,\beta^{\alpha}}{w^{\alpha+1}} \qquad \alpha>0,\ w>\beta>0 \tag{4.2}$$

with survival function

$$\Pr[W>w|\alpha,\beta] = \min\left[\left(\frac{\beta}{w}\right)^{\alpha},\ 1\right] \tag{4.3}$$

Now we recall the well known fact, setting $\beta=e^{\gamma}$, $\log W=X$, $\log w=x$, that

$$\Pr[W>w|\alpha,\beta] = \Pr[X>x|\alpha,e^{\gamma}] = \min[e^{-\alpha(x-\gamma)},\ 1]=\Theta. \tag{4.4}$$

The latter is equivalent to (2.3), the translated exponential survival function, and thus Θ is invariant.

If, for the pareto parameters (α,β), we employ

$$p(\alpha,\beta)\propto(\alpha\beta)^{-1}, \tag{4.5}$$

then the transformation $\beta=e^{\gamma}$ yields

$$p(\alpha,\gamma)\propto\alpha^{-1} \tag{4.6}$$

which is identical to (2.6).

Hence all results presented for the translated exponential case regarding Θ and $\bar{Y}$ can now be utilized for Pareto distributed variables by making the following simple transformations

$$X_i = \log W_i \qquad i = 1,\dots,N$$

$$\bar{x} = N^{-1} \sum_{i=1}^{N} \log w_i = \log(w_1,\dots,w_N)^{1/N} = \log G$$

$$m = \min(\log w_i, \dots, \log w_N) = \log \min(w_1,\dots,w_N)$$

Recently there has been a good deal of analysis made on Bayesian estimation of the pareto parameters α and β by Arnold and Press (1982). They generally restrict themselves to parametric estimation but use general prior distributions of which the "vague" prior used here is a special case. They point out, among other things, that the vague prior distribution for β results in its marginal posterior density being unbounded in the neighborhood of $\beta = 0$, and that this characteristic is not eliminated as the sample size grows.

They regard it as particularly distressing that the marginal distribution of β should possess such a property and suggest ways of avoiding this. But they do not indicate reasons for their discomfiture. The calculation of the marginal posterior density of γ from the translated exponential is easily obtained as

$$p(\gamma|\underset{\sim}{x}) = \frac{(d-1)(\bar{x}-m)^{d-1}}{(\bar{x}-\gamma)^d} \qquad \gamma < m,\ d > 1 \tag{4.7}$$

and we note that no such problem exists since $m < \bar{x}$. In fact the posterior mode is bounded at $\gamma = m$ and the density monotonically declines as γ decreases - which certainly is a reasonable posterior density. However, if we consider the posterior density of the transformed variable $\beta = e^{\gamma}$, then setting $x_i = \log w_i$ so that in (4.4) $\bar{x}$ is now replaced by log G where G is the geometric mean of the w_i's and m the minimum of the x's is replaced by $\log m'$ where m' is the minimum of the w_i's, we obtain

$$p(\beta|\underset{\sim}{w}) = \frac{(d-1)\left(\log \frac{G}{m'}\right)^{d-1}}{\beta\left(\log \frac{G}{\beta}\right)^{d}} \qquad \beta < m' \tag{4.8}$$

Here a second mode at $\beta = 0$ is introduced, but we see that this is just a result of the transformation and should not cause any particular distress. Further, from the point of view of prediction there is no intrinsic interest in making inferential statements about either β or α so anomalies, if any, that do not effect the prediction of $\bar{Y}$ should cause us little or no concern.

The work presented here depends on a simplifying but realistic assumption that the minimum observation is fully observed and that there is at least one other non-censored value. Although (2.24) is still a probability for d=1, its dependence on z reflects only whether or not z exceeds the minimum observation. This is unsatisfactory but hardly surprizing because the existence of the posterior distribution of α requires $d \geq 2$ if γ is unknown. When γ is known then we define $x - \gamma = v$ and we can use the results in Geisser (1982). Here $d \geq 1$ is necessary for reasonable results. When all the observations are censored (or all but one in the unknown γ case) then prior knowledge must convey more information than we have expressed here in order to obtain results. For example, the prior distributions of Arnold and Press (1982) or Lwin (1972) could be utilized.

5. An Example

Dyer (1981) reported annual wage data (in multiples of 100 dollars) of a random sample of 30 production-line workers in a large industrial firm as follows:

112, 154, 119, 108, 112, 156, 123, 103, 115, 107,
125, 119, 128, 132, 107, 151, 103, 104, 116, 140,
108, 105, 158, 104, 119, 111, 101, 157, 112, 115.

He determined that the Pareto distribution provided an adequate fit for the data. The following calculations yield in dollars

$$m' = 10{,}100, \qquad \log m' = m = 9.2203,$$
$$G = 11{,}959, \qquad \log G = \bar{x} = 9.3893.$$

Suppose we were interested in calculating the probability that no more than 10% of the next M randomly selected production-line workers had an annual wage larger than 15,000 dollars.

In Table 1 we present exact and approximate values for a series of values of M

Table 1:

Exact and Approximate values of $\Pr[\bar{Y} \leq .1 | w = 15{,}000]$.

M	Exact	F-Approximation	χ^2-Approximation
10	.6998	.7064	.8048
20	.6294	.6205	.6502
30	.5939	.5776	.5848
40	.5720	.5514	.5498
50	.5569	.5335	.5282
60	.5457	.5206	.5135
70	.5372	.5107	.5029
80	.5304	.5030	.4949
90	.5248	.4967	.4887
100	.5202	.4916	.4836
∞	.4378	.4378	.4378

While the F-approximation appears for the most part to dominate the χ^2-approximation we note that even for $M = 100$ it still induces a relative error of approximately 5% indicating that a search for a better approximation may be appropriate. Also we note that the asymptotic value $(M = \infty)$ is about 16% smaller than the exact value for an M as large as 100.

6. Further Remarks

This work is easily extended to an appropriate natural conjugate prior family for (α,γ), where we define known hyperparameters $d_0, N_0, \bar{x}_0, m_0$. Then $\Pr(\bar{y} = \frac{r}{M} \mid z)$ is of the same form as in (2.24) but substituting for d, N, $\bar{x}$ and m,

$$
\begin{aligned}
d^* &= d_0 + d, \\
N^* &= N_0 + N \\
\bar{x}^* &= (N_0\bar{x}_0 + N\bar{x})/(N_0 + N) \\
m^* &= \min(m_0, m).
\end{aligned}
\tag{6.1}
$$

respectively. Use of such a prior distribution assumes, of course, that the information available can be conveniently subsumed in the form of a sample similar in nature to the one to be collected.

As noted in section 2, many of the calculations are greatly simplified by assuming the minimum was fully observed. This will clearly be the case in many well controlled experiments where censoring is generally applied as a device to terminate an experiment for economic or other reasons. In certain medical and other less well controlled studies, there may be dropouts (say unrelated to the administered agents) prior to any failure. However, ignoring them generally has minimal influence on inferences involving future observables. The reason is that little information inheres to these early censored observations in the sense that their effect is confined to regions of relatively low density.

I am indebted to S. James Press for bringing to my attention his own work with Arnold and also Dyer's paper.

References

Arnold, B.C. and Press, S.J. (1982). "Bayesian inference for Pareto populations," To appear in Journal of Econometrics.

de Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives, Ann. Inst. H. Poincaré, 7, 1-68.

Dyer, D. (1981). "Structural probability bounds for the strong Pareto law, Canadian Journal of Statistics, 9, 71-77.

Fraser, D.A.S. (1979). Inference and Linear Models, McGraw-Hill, New York.

Geisser, S. (1982). Aspects of the predictive and estimative approaches in the determination of probabilities, Biometrics (supplement), 38, 1, 75-93.

Lwin, T. (1972). "Estimation of the tail of Paretian law, Skand. Aktuar., 55, 170-178.